Blood First Edition Paper, prepublished online July 20, 2006; DOI 10.1182/blood-2005-01-026690

Risk of Non-Hodgkin lymphoma (NHL) in relation to germline variation in DNA repair and related genes.

Running Head: NHL risk and DNA repair gene variants

Deirdre A. Hill[1], Sophia S. Wang[2], James R. Cerhan[3], Scott Davis[4], Wendy Cozen[5], Richard K. Severson[6], Patricia Hartge[2], Sholom Wacholder[2], Meredith Yeager[7], Stephen J. Chanock[7], Nathaniel Rothman[2].

[1]*University of New Mexico, Albuquerque, Cancer Center and Department of Internal Medicine, Albuquerque, NM, USA.*. [2]*Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, MD, USA.* [3]*Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, MN, USA and University of Iowa, Iowa City, IA, USA.* [4]*Fred Hutchinson Cancer Research Center and the University of Washington, Seattle, WA, USA.* [5]*Dept. Of Preventive Medicine, University of Southern California Keck School of Medicine, Los Angeles, CA, USA.* [6]*Karmanos Cancer Institute and Dept. of Family Medicine, Wayne State University, Detroit, MI, USA.* [7]*Core Genotyping Facility, Advanced Technology Corporation, National Cancer Institute, NIH, DHHS, Gaithersburg, MD, USA.*

This study was supported by Public Health Service contracts N01-PC-65064, N01-PC-67008, N01-PC-67009, N01-PC-67010, and N02-PC-71105.

Correspondence should be addressed to:

Deirdre A. Hill, Ph.D., Dept. of Internal Medicine, UNM School of Medicine, 1 University of New Mexico, MSC 10 5550, Albuquerque, NM, 87131-0001. Telephone: (505) 272-5049. Fax: (505) 272-8572. Email: dahill@salud.unm.edu.

Word Count: 3693

Scientific Heading: Neoplasia

Author Contributions: This project was conceived and led by the NCI-SEER NHL group: P. Hartge, N. Rothman, S. Wang, S. Chanock, R. Severson, S. Davis, W. Cozen, and J. Cerhan. Genotyping, quality control specimens, and bioinformatics support was provided by M. Yeager-Jeffrey and S. Chanock. D. Hill selected the genes and variants to be assayed, with the assistance of S. Wang, N. Rothman and S. Chanock. The statistical analysis was conducted by D. Hill with input from S. Wang, N. Rothman, and P. Hartge, and from S. Wacholder on methods for controlling the chance that a reported finding of association is a true positive. The manuscript was drafted by D. Hill and was revised with contributions from all coauthors. All authors reviewed and approved the manuscript.

Abstract

Chromosomal translocations, insertions, and deletions are common early events in NHL carcinogenesis, and implicated in their formation are endogenous processes involved in antigen-receptor diversification, such as V(D)J recombination. DNA repair genes respond to the double- and single-strand breaks induced by these processes, and may influence NHL etiology. We examined 34 genetic variants in 19 genes within or related to five DNA repair pathways among 1172 cases and 982 matched controls who participated in a population-based NHL study in Los Angeles, Seattle, Detroit and Iowa from 1998-2000. Cases were more likely than controls to have the *RAG1* 820 Arg/Arg (Odds ratio (OR) 2.7; 95% confidence interval (CI) 1.4-5.0) than Lys/Lys genotypes, with evidence of a gene dosage effect (p-trend=0.0008) and less likely to have the *LIG4* (DNA Ligase IV) 9 Ile/Ile (OR 0.5; 95% CI 0.3-0.9) than Thr/Thr genotype (p-trend =0.03) in the non-homologous end joining (NHEJ)/V(D)J pathway. These NHEJ/V(D)J-related gene variants represent promising candidates for further studies of NHL etiology, and require replication in other studies.

Introduction

Chromosomal translocations are a hallmark of non-Hodgkin lymphoma (NHL) and can arise as a consequence of misrepair of DNA double strand breaks. Major translocations identified in NHL include those fusing bcl2 with immunoglobulin (Ig) H in approximately 80% of follicular lymphomas; [1] bcl2, bcl6, or myc with IgH in approximately 50% of diffuse large B-cell lymphomas (DLBCL); [2] and myc with one of several Ig loci in 80% or more of Burkitt lymphomas. [1] These and other genomic rearrangements (e.g., insertions, deletions) are thought to occur early in malignant transformation in most NHLs. Accordingly, unrepaired or misrepaired DNA strand breaks could be critical events in lymphomagenesis.

While some translocations are generated by aberrant repair of double-strand breaks induced by ionizing radiation or other external exposures, alterations in endogenous processes such as V(D)J recombination could also contribute to such rearrangements. V(D)J recombination involves the deliberate introduction of double-strand breaks that reshuffle dozens of Ig building blocks, the V, D, and J segments. This process produces a highly diverse repertoire of antibodies, which are induced by a wide spectrum of antigenic challenges. Errors by the DNA repair genes responsible for ligating the V, D, and J segments, the non-homologous end-joining (NHEJ) genes, are implicated at the sites of rearrangements characteristic of NHL [3,4] In addition, two steps that follow V(D)J in B-cell maturation, class-switch recombination and somatic hypermutation, also introduce DNA strand breaks.[5] The observation of NHL-associated translocations[6] or aberrant hypermutation[7] preferentially involving those regions suggest that misrepair of DNA breaks during these events could also contribute to lymphomagenesis. The

fidelity of repair of such breaks and other types of DNA damage implicated in lymphoma development may directly or indirectly involve any of five overlapping DNA repair pathways: (i) NHEJ, (ii) homologous recombination repair, (iii) nucleotide excision repair, (iv) base excision repair, and (v) direct damage reversal.

NHEJ, one of two major double-strand break repair pathways, is considered "error-prone" in part because it does not employ a homologous strand as a template to repair DNA breaks, but instead allows ligation by introducing small nucleotide insertions or deletions into DNA. NHEJ has been shown in vitro to produce translocations.[8] Mice deficient in any of several NHEJ genes on a p53-/- background are predisposed to develop immunodeficiency and pro- B-cell lymphomas,[9,10] and the lymphomas demonstrate myc-IgH translocations reminiscent of those in human Burkitt lymphoma. Mutations in the NHEJ genes *RAG* or *DCLRE1C* (known as Artemis) in humans lead to severe combined immunodeficiency (SCID) syndrome, which often involves almost complete abrogation of B- and T-cells.[11] However, two out of four carriers of a *DCLRE1C* mutation that allowed partial B- and T-cell expression developed lymphomas.[12] Thus, accumulating evidence suggests that NHEJ/V(D)J genes may participate in a vital way in lymphomagenesis.

Homologous recombination (HR) repair, the second major double-strand break repair pathway, has been described as "error-free" but can also result in translocations.[13] Inherited mutations in HR genes have been recognized in familial cancer syndromes that involve an elevated lymphoma risk, including Bloom syndrome (BLM),[14] Fanconi anemia (FA),[15] and Nijmegen Breakage syndrome (NBS).[16] Development of NHL, while possibly facilitated by

NHEJ or HR genes, may also be influenced indirectly by several genes active in base excision repair (BER), [17,18] a DNA single-strand break repair pathway, or nucleotide excision repair (NER), a pathway involved in repair of DNA damage induced by ultraviolet radiation[19,20] or bulky adducts. Finally, MGMT, a gene which participates in "direct reversal" of DNA damage by removal of O6-methylguanine adducts, is frequently hypermethylated in NHL tumors,[21,22] and is also hypothesized to contribute to lymphoma development.

As several lines of evidence support the involvement of DNA repair and related genes in the etiology of NHL, particularly in the double-strand break pathway, we selected 19 genes that play an important role in or are related to five DNA repair pathways for analysis in a large, population-based case-control study of NHL in the United States. Here, we report results for a total of 34 genetic variants, which were selected based on theoretical or experimental evidence of functionality and previous evidence of association.

Methods

Study Population

A detailed description of the study methods for the National Cancer Institute (NCI) Surveillance, Epidemiology, and End Results (SEER) case-control study of NHL has previously been published.[23] Briefly, individuals aged 20-74 years diagnosed with incident NHL from July 1, 1998 to June 30, 2000 were identified in four U.S. SEER population-based cancer registries: the metropolitan areas of Detroit, Los Angeles, and Seattle, and the state of Iowa. Eligible controls were selected from the general population in the four registry areas, using random-digit dialing (age 20-64 years) or Medicare eligibility files (age 65-74 years), and were frequency-matched to NHL cases by age (5-year intervals), sex, race (White, African-American, Asian, other) and SEER study site. Eligible cases or controls who were identified by themselves or a physician as HIV-infected were excluded.

Of the 2248 selected eligible NHL cases, 320 (14.2%) died before interview, 57 (2.5%) were not interviewed because of physician refusal, and 143 (6.4%) were unable to be located. The remaining 1728 were contacted, of whom 274 (15.9%) refused to participate and 133 (7.7%) were not interviewed because of nonresponse, illness, impairment, or other reasons. Thus, 1321 eligible cases were interviewed, for a 76% participation rate and a 59% response rate. Of 2409 eligible controls, 311 (13%) could not be located, 28 (1%) died before interview, and 24 (1%) moved out of the area. The remaining 2046 were contacted, and of these, 839 (41%) declined to participate and 150 (6%) were not interviewed for other reasons, yielding 1057 eligible controls, and a 52% participation and a 44% response rate. Participants signed an informed consent form, received a computer-assisted personal interview regarding known or suspected NHL risk factors,

and donated a blood (773 cases, 668 controls) or buccal cell (399 cases, 314 controls) sample. Of the 1172 case and 982 control participants, the study population included in the genetic analyses consists of the 1150 cases and 956 controls for whom DNA could be extracted and subsequently genotyped for polymorphisms in DNA repair and related genes. The study was reviewed and approved by institutional review boards at the NCI and at each of the SEER study sites.

Laboratory Methods

DNA extraction. DNA was extracted from blood clots or buffy coats at BBI Biotech Repository (Gaithersburg, MD) using Puregene Autopure DNA extraction kits (Gentra Systems, Minneapolis, MN). Phenol chloroform extraction methods were used to obtain DNA from buccal cell samples collected via mouthwash. DNA was stored at + 4 C° until genotyping.

Genotyping: All genotyping was conducted at the NCI Core Genotyping Facility (Advanced Technology Corporation, Gaithersburg, MD) using either the Taqman, Sequenom, or MGB Eclipse sequencing platforms. Assays used to examine gene variation were developed and validated using previously published procedures.[24] Details regarding platforms, primers, and assays conditions can be obtained at http://snp500cancer.nci.nih.gov. All laboratory personnel were blinded as to the case or control status of samples. The frequency of "undetermined" genotype or "no PCR amplification observed" (generally 2%-4% total) for any particular genotype did not differ among cases and controls. After DNA derived from blood samples had been genotyped, a preliminary analysis was conducted, and only those variants which demonstrated a relationship with NHL risk were genotyped using the buccal cell DNA (due to limited DNA yield from those specimens). Less than 1% of the 140 included blinded samples (40

blood donor replicates; 100 duplicates from study subjects) were discordant for each genotype. In particular, no homozygous wild type genotype for any included gene was classified as homozygous variant, or vice-versa. For each genotype, 4 samples of known homozygous wild-type, heterozygote, or homozygous variant genotypes also were included for quality control in each plate of 386 samples, as well as 4 DNA negative controls.

Statistical Methods

We examined whether the distribution of genotypes in controls was consistent with Hardy-Weinberg equilibrium (HWE) using the X^2 test. All control genotype frequencies were in accord with HWE in each race/ethnic group. We conducted unconditional logistic regression to calculate odds ratios (OR) and 95% confidence intervals (95% CI) for the relationship between genotype and NHL risk, adjusting for the matching factors of age (<55, 55-64, $\geq$ 65), race (white, African-American, Asian/Other), sex, and study site. Odds ratio estimates were determined using the common homozygote genotype as the referent group. We also examined whether the relationship between gene variants and NHL risk differed by race (White, African-American), sex, age (< age 60 years, $\geq$ age 60 years), and REAL/WHO tumor pathology group (follicular, DLBCL, T-cell, other, unknown). Risk estimate heterogeneity between tumor groups was tested by designating one tumor pathology group as "cases" and another as "controls". Statistical interaction on a multiplicative scale between genotypes, or between a particular genotype and race, sex or age, was assessed by including main effects terms for each variable in the logistic regression model, and adding a product term (gene1 * gene2 or gene1* sex). All analyses were conducted using SAS software (SAS version 8, Cary, N.C.)

There is a lack of consensus regarding the optimal approach to address the false positive probability of SNP associations. We therefore evaluated the robustness of our results using two complementary methods - the False Discovery Rate (FDR)[25] and the False Positive Report Probability (FPRP).[26] FDR is the expected ratio of erroneous rejections of the null hypothesis to the total number of rejected hypotheses among the SNPs analyzed in this report. We applied the FDR method to the p-value for trend as this allows for the fewest number of comparisons and thus degrees of freedom, and to assess the additive model. We applied the FPRP method, which controls the probability that a single SNP association is a false positive report, for a range of prior probabilities (i.e., 0.001 to 0.1) that the given SNP is truly associated with risk of NHL. The same prior range was used in a previous large pooled report of cytokine polymorphisms and NHL[27] and reflects the extent to which a candidate SNP is likely to be functional and located in a gene that plays a role in the pathogenesis of NHL. We used an FPRP criteria of 0.20 (recommended in the original presentation of the method) [26] to identify which, if any, findings should be considered noteworthy.

Results

The median age of NHL cases and controls was 58 and 61, respectively (Table 1). As expected due to matching, cases and controls were broadly similar in sex and race, with the exception of a higher proportion of African-American controls (14%) than cases (8%). Thirty-four genetic variations among 19 genes in or related to five DNA repair pathways were examined in the present study (Table 2).

Non-homologous end joining and V(D)J recombination genes

The *LIG4* (DNA Ligase IV) 9 Ile variant allele was less common among NHL cases than controls overall (Thr/Ile OR=0.9; 95% CI=0.7-1.1; Ile/Ile OR=0.5; 95% CI=0.3-0.9; p-trend=0.03) (Table 3). The reduced risk was also apparent for follicular lymphoma (Ile/Thr or Ile/Ile OR 0.7; 95% CI 0.5-1.0) and DLBCL (Ile/Thr or Ile/Ile OR 0.8; 95% CI 0.6-1.0) (Table 4). NHL cases were more likely than controls to have the *RAG1* 820 Arg variant allele (Lys/Arg OR=1.3; 95% CI=1.0-1.6; Arg/Arg OR=2.7; 95% CI=1.4-5.0; p-trend=0.001) (Table 3). When examined among NHL pathology groups, follicular lymphoma cases were also more likely than controls to have inherited this variant (Lys/Arg OR=1.3; 95% CI= 0.9-1.8; Arg/Arg OR=5.1; 95% CI 2.3-11.7; p-trend=0.0009) (Table 4). The relationship with NHL risk was not strong among other subtypes, however the difference between those subtypes and follicular was not significant. Other NHEJ gene variants were not associated with altered NHL risk (Supplementary Tables 1 and 2). Among those with both a *RAG1* 820 Arg and *LIG4* 9 Thr (increased risk) allele, risk of NHL was not elevated beyond that expected from the joint multiplicative effects of the two risk factors (data not shown). Risk among individuals with one

or more *RAG1* or *LIG4* variants also did not vary by race, sex, or age (< 60, ≥ 60) (data not shown).

Homologous recombination (HR) repair genes

NHL risk was examined in relation to 14 variants in six homologous recombination repair genes: *BRCA2, NBS1, TP53, WRN, XRCC2,* and *XRCC3*. Overall, NHL cases were 1.5-fold more likely than controls to be homozygous for the *BRCA2* 372 His/His genotype (95% CI 1.0-2.1) (Table 3). Although risks of follicular and DLBC lymphoma were similarly elevated only among homozygotes, risk of T-cell lymphoma increased with an increasing number of *BRCA2* 372 His alleles (1.8-fold among *BRCA2* Asp/His heterozygotes and 3.0-fold among His/His homozygotes (p-trend=0.003) (Table 4). The *WRN* V114I variant was less common among cases than controls overall, and NHL risk decreased with an increasing number of *WRN* 114 Ile alleles (p-trend .04) (Table 3). The reduced risk was not confined to a particular NHL pathology group (Table 4). The altered NHL risks among individuals with *BRCA2* or *WRN* variants were equally apparent among participant subgroups defined by race, sex, or age (data not shown). Individuals with other HR variants did not have an altered NHL risk (Supplementary Tables 1 and 2).

Nucleotide Excision Repair (NER) genes

Seven variants in six NER genes (*ERCC1, ERCC2, ERCC4, ERCC5, XPC,* and *RAD23B*) were examined in relation to NHL risk. Overall, NHL cases were no more likely than controls to have any NER variant, and apparent differences in participant subgroups defined by race, sex, or age were ascribable to chance (Supplementary Tables 1 and 2, and data not shown).

Base Excision Repair (BER) genes

Risk of NHL was examined in relation to five variants in three BER genes (*PARP*, *APEX1*, and *XRCC1*). The *XRCC1* 194 Trp allele was associated with a moderately increased risk (1.4-fold) of NHL overall (Table 3), but did not differ significantly by NHL pathology group (Table 4), race, sex, or age (data not shown). The presence of other BER variants was not related to an altered NHL risk (Supplementary Tables 1 and 2).

Direct Reversal of Damage

Three variants in *MGMT*, a gene active in direct reversal of DNA damage, were examined and none were associated with NHL risk overall (Supplementary Tables 1 and 2).

Discussion

Our results suggest that variants in several DNA repair or V(D)J pathway genes may be related to an altered risk for NHL or its subtypes. Particularly, homozygotes for the *RAG1* 820 Arg missense substitution had a 2.7-fold increased risk for NHL, with evidence of a gene-dosage effect. While *RAG1* is not considered a DNA repair gene, it participates in V(D)J recombination with genes active in NHEJ repair. Functional studies of the *RAG1* 820 Arg variant are lacking, although assessment of functional effects using the SIFT (Sorting Intolerant From Tolerant (http://blocks.fhcrc.org/sift) program indicated that the polymorphism was likely to be "not tolerated" (probability < .01). Individuals with highly penetrant, disruptive *RAG1* mutations are immunodeficient, have partial (Omenn syndrome) or virtually absent (SCID) V(D)J activity[29] and experience severe B- and T-cell defects.[11] The RAG1 and RAG2 core protein complex has been shown *in vitro* to cleave DNA at specific sites and insert the cleaved segment at target sites unrelated to V(D)J, creating a translocation.[8] The biological plausibility of *RAG1* involvement in events that initiate translocations, the deleterious nature of the substitution, and the dose-response found in this study support the possibility that individuals with the *RAG1* 820 Arg variant may have an altered risk of lymphoma.

Functional and epidemiological studies also support our finding that the *LIG4* T9I polymorphism may be related to altered lymphoma risk. When evaluated for functional effects using SIFT, the T9I substitution was predicted to be "not tolerated" (probability = .01). In addition, the *LIG4* T9I variant construct, when expressed in cell culture with the A3V variant in linkage disequilibrium, demonstrated 2 - 3-fold lower DNA double-strand break ligation activity and 2-fold lower adenylation activity than wild type *LIG4*.[30] Our observation that individuals

with the T9I variant have a decreased lymphoma risk is consistent with the 3-fold reduced risk of lymphoma, and the 5-fold decreased risk of multiple myeloma (n=7 and 4 cases, respectively), among *LIG4* 9 Ile/Ile homozygotes in a previous case-control study.[31] Although an XRCC4-DNA Ligase IV complex ligates DNA ends during DSB repair, no interaction on a multiplicative scale was observed between polymorphisms in the two genes. Also, lymphoma risk among those with one or more NHEJ gene variants was not modified by *TP53* genotype (data not shown), although *TP53* status appears to alter NHEJ gene response in mice.[9] Homozygotes for the *RAG1* or *DNA Ligase IV* variants had the strongest alterations in NHL risk in this study, although such individuals are rare (1.6% and 3.0% of controls, respectively). However, if heterozygotes for these variants have an altered NHL risk comparable to that suggested by our data, then their higher prevalence (20.0% and 29.5%) implies that a greater proportion of NHL risk in the population would be attributable to heterozygosity.

Although NHL risk was most strongly related to variants in the V(D)J/NHEJ pathway, polymorphisms in two genes (*BRCA2, WRN*) involved in double-strand break resolution via HR (a pathway also known to induce translocations[13]) were also related to an altered NHL risk. In some[32-34] but not all[35,36] previous studies, *BRCA2* 372 His/His homozygotes have had an approximately 1.4-fold increased risk of breast or ovarian cancer, but their NHL risk has not previously been evaluated. In our study, homozygotes also had a 1.4-fold elevated risk of all lymphoma and of follicular and DLBC lymphoma, while T-cell lymphoma risk was elevated among individuals with the His variant, with evidence of an effect of gene-dosage. In a previous study, risk of lymphoma was nonsignificantly increased (OR=1.8) in relatives of *BRCA2* mutation carriers.[37] We also found that individuals with a *WRN* 114 Ile allele had a dose-

dependent reduced risk of NHL that was not confined to any tumor type, sex, or age-specific subgroup. However, prior studies have not evaluated this allelic change, and the V114I substitution is predicted to be tolerated by the SIFT program. Individuals carrying mutations in *WRN*, which predisposes to Werner Syndrome, have an increased risk of sarcomas, melanomas, and thyroid cancer, and one leukemia but no lymphomas have been reported among 124 individuals.[38] Although *PARP* interacts with *WRN* in DNA repair processes,[17] *PARP* genotype status did not alter the decreased NHL risk among those with at least one *WRN* 114I variant allele (data not shown).

The involvement of *XRCC1* and other BER genes in the processing of immunoglobulin rearrangement intermediates during somatic hypermutation and class-switch recombination[18] argues that BER genes could participate in early events in lymphomagenesis. In the overall analysis, individuals with *XRCC1* R194W variant alleles had a moderately increased risk of NHL, and this finding was not limited to any specific tumor-type, sex, or age group. Although this variant has been related to risk of tumors at other sites, [39,40] in a recent study, individuals who inherited one or more R194W variant alleles did not have an altered risk of follicular lymphoma.[41]

Our results should be considered in the context of the strengths and limitations of the study, as well as the possibility that some findings are false positives, given the number of relationships examined. Strengths of this study include the population-based design and the large sample size. In addition, excellent laboratory quality control measures, including a concordance of 99% or greater for replicate genotypes in blinded samples, testify to the

reliability of the data. However, while similar to those in many recent case-control studies, the response rates for this study were lower than desirable for both cases and controls. Participation among cases and controls was unlikely to be differential by genotype, and variant prevalence in white non-Hispanic study controls corresponded closely to that observed in random samples of similar individuals (http://snp500cancer.nci.nih.gov).[24] However, if any included polymorphism is related to early mortality from NHL, the prevalence in participating cases could be altered, introducing bias in odds ratio estimates. In addition, although the sample size provided sufficient power to evaluate main effects of low frequency variants, the study did not have ample power to evaluate most gene-gene interactions or to determine whether there were statistically significant differences between NHL histologies. Because survival and genetic alterations in NHL tumors differ by histology, we believe inclusion of histology information is helpful to examine subgroup heterogeneity, however, we did not observe major heterogeneity between subgroups. We presented results for all study subjects, adjusting for race, but our key findings, including the *LIG4 T91* and *RAG1 K820R* associations, remained statistically significant and were essentially identical in magnitude when analyses were restricted to White non-Hispanics (data not shown).

An assessment of the probability that a statistically significant result at $p < 0.05$ is a false positive finding can aid in the interpretation of study findings. We evaluated our results using the FDR[25] and FPRP[26] approaches, as described in the Methods. The FDR value of the *RAG1* K820R variant was 0.02, taking into account all SNPS tested for association with risk of NHL overall in this report, and the FPRP value was below our criterion of 0.2 (for a prior probability of association of 0.01 or higher, expected odds ratio of 1.3 or higher, and observed odds ratio from the additive model for all NHL: OR: 1.37, 95% CI: 1.14 - 1.65, p trend = 0.0008); that is,

both methods indicate that that the association may be particularly robust, and suggest that there is only a small chance that the *RAG1* K820R finding is a false positive. Although no other findings were deemed noteworthy after carrying out FDR and FPRP calculations, exploration of these associations in larger studies with greater power may be of value, particularly using tagged SNPs to obtain full genomic coverage of the most promising candidate genes.

In summary, our results suggest that inherited variants in NHEJ or V(D)J genes may alter risk of NHL, but our findings require replication by other studies, with an eventual goal of pooling across multiple investigations to evaluate the robustness of the findings. Investigation of the phenotypic relevance of the identified genetic variation, the contribution of other genes in the pathway, and potential interactions with other risk factors may ultimately yield new insights into the poorly understood process of lymphomagenesis.

Acknowledgements: We would like to gratefully acknowledge all staff and scientists at each of the SEER registry sites for the conduct of the study's field effort and collection of biological specimens. We also are grateful to Robert Welch and Sunita Yadavalli at the NCI Core Genotyping Facility for their careful handling of specimens and meticulous analysis of genotyping data.

REFERENCES

1. Fan YS, Rizkalla K. Comprehensive cytogenetic analysis including multicolor spectral karyotyping and interphase fluorescence in-situ hybridization in lymphoma diagnosis. A summary of 154 cases. Cancer Genet Cytogenet. 2003;143:73-79.

2. Yano T, van Krieken JH, Magrath IT, Longo DL, Jaffe ES, Raffeld M. Histogenetic correlations between subcategories of small noncleaved cell lymphomas. Blood. 1992;79:1282-1290.

3. Tsujimoto Y, Jaffe E, Cossman J, Gorham J, Nowell PC, Croce CM. Clustering of breakpoints on chromosome 11 in human B-cell neoplasms with the t(11;14) chromosome translocation. Nature. 1985;315:340-343.

4. Jager U, Bocskor S, Le T, et al. Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: novel insights into the mechanism of t(14;18) translocation. Blood. 2000;95:3520-3529.

5. Papavasiliou FN, Schatz DG. Cell-cycle-regulated DNA double-stranded breaks in somatic hypermutation of immunoglobulin genes. Nature. 2000;408:216-221.

6. Shiramizu B, Magrath I. Localization of breakpoints by polymerase chain reactions in Burkitt's lymphoma with 8;14 translocations. Blood. 1990;75:1848-1852.

7. Pasqualucci L, Neumeister P, Goossens T, et al. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature. 2001;412:341-346.

8. Hiom K, Melek M, Gellert M. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. Cell. 1998;94:463-470.

9. Gao Y, Ferguson DO, Xie W, et al. Interplay of p53 and DNA-repair protein XRCC4 in tumorigenesis, genomic stability and development. Nature. 2000;404:897-900.

10. Difilippantonio MJ, Petersen S, Chen HT, et al. Evidence for replicative repair of DNA double-strand breaks leading to oncogenic translocation and gene amplification. J Exp Med. 2002;196:469-480.

11. Villa A, Sobacchi C, Notarangelo LD, et al. V(D)J recombination defects in lymphocytes due to RAG mutations: severe immunodeficiency with a spectrum of clinical presentations. Blood. 2001;97:81-88.

12. Moshous D, Pannetier C, Chasseval Rd, et al. Partial T and B lymphocyte immunodeficiency and predisposition to lymphoma in patients with hypomorphic mutations in Artemis. J Clin Invest. 2003;111:381-387.

13. Richardson C, Jasin M. Frequent chromosomal translocations induced by DNA double-strand breaks. Nature. 2000;405:697-700.

14. German J, Ellis NA. Bloom Syndrome. *In:* The Genetic Basis of Human Cancers. Vogelstein B, Kinzler K, eds. New York, McGraw-Hill 1998. pgs 301-315.

15. Auerbach AD, Buchwald M, Joenje H. Fanconi Anemia. *In:* The Genetic Basis of Human Cancers. Vogelstein B, Kinzler K, eds. New York, McGraw-Hill 1998. pgs. 317-332.

16. International Nijmegen Breakage Syndrome Study Group. Nijmegen breakage syndrome. Arch Dis Child. 2000;82:400-406.

17. Morrison C, Smith GC, Stingl L, et al. Genetic interaction between PARP and DNA-PK in V(D)J recombination and tumorigenesis. Nat Genet. 1997;17:479-482.

18. Akbari M, Otterlei M, Pena-Diaz J, et al. Repair of U/G and U/A in DNA by UNG2-associated repair complexes takes place predominantly by short-patch repair both in proliferating and growth-arrested cells. Nucleic Acids Res. 2004;32:5486-5498.

19. Smedby KE, Hjalgrim H, Melbye M, Torrang A, Rostgaard K, Munksgaard L, Adami J, Hansen M, Porwit-MacDonald A, Jensen BA, Roos G, Pedersen BB, Sundstrom C, Glimelius B, Adami HO. Ultraviolet radiation exposure and risk of malignant lymphomas. J Natl Cancer Inst. 2005;97:199-209.

20. Adami J, Frisch M, Yuen J, Glimelius B, Melbye M. Evidence of an association between non-Hodgkin's lymphoma and skin cancer. BMJ. 1995;310:1491-1495.

21.Esteller M, Hamilton SR, Burger PC, Baylin SB, Herman JG. Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is a common event in primary human neoplasia. Cancer Res. 1999;59:793-797.

22. Esteller M, Gaidano G, Goodman SN, ET AL. Hypermethylation of the DNA repair gene O(6)-methylguanine DNA methyltransferase and survival of patients with diffuse large B-cell lymphoma. J Natl Cancer Inst. 2002;94:26-32.

23. Chatterjee N, Hartge P, Cerhan JR, et al. Risk of non-Hodgkin's lymphoma and family history of lymphatic, hematologic, and other cancers. Cancer Epidemiol Biomarkers Prev. 2004;13:1415-1421.

24. Packer BR, Yeager M, Burdett L, et al.. CancerSNPs (formerly known as SNP500Cancer): A Public Resource for Sequence Validation, Assay Development, and Frequency Analysis for Genetic Variation in Candidate Genes. Nucleic Acids Res. 2006; 34 :D617-D621.

25. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society B. 1995;57:289-300.

26. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst. 2004;96:434-442.

27. Rothman N, Skibola CF, Wang SS, et al. Genetic variation in TNF and IL10 and risk of non-Hodgkin lymphoma: a report from the InterLymph Consortium.Lancet Oncol. 2006 ;7:27-38.

28. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. Lancet. 2003; 361: 865-872.

29. Corneo B, Moshous D, Gungor T, et al. Identical mutations in RAG1 or RAG2 genes leading to defective V(D)J recombinase activity can cause either T-B-severe combined immune deficiency or Omenn syndrome. Blood. 2001;97:2772-2776.

30. Girard PM, Kysela B, Harer CJ, Doherty AJ, Jeggo PA. Analysis of DNA ligase IV mutations found in LIG4 syndrome patients: the impact of two linked polymorphisms. Hum Mol Genet. 2004;13:2369-2376.

31. Roddam PL, Rollinson S, O'Driscoll M, Jeggo PA, Jack A, Morgan GJ. Genetic variants of NHEJ DNA ligase IV can affect the risk of developing multiple myeloma, a tumour characterised by aberrant class switch recombination. J Med Genet. 2002;39:900-905.

32. Healey CS, Dunning AM, Teare MD, et al. A common variant in BRCA2 is associated with both breast cancer risk and prenatal viability. Nat Genet. 2000;26:362-364.

33. Spurdle AB, Hopper JL, Chen X, et al. The BRCA2 372 HH genotype is associated with risk of breast cancer in Australian women under age 60 years. Cancer Epidemiol Biomarkers Prev. 2002;11:413-416.

34. Auranen A, Spurdle AB, Chen X, et al. BRCA2 Arg372His polymorphism and epithelial ovarian cancer risk. Int J Cancer. 2003;103:427-430.

35. Wenham RM, Schildkraut JM, McLean K, et al. Polymorphisms in BRCA1 and BRCA2 and risk of epithelial ovarian cancer. Clin Cancer Res. 2003;9:4396-4403.

36. Cox DG, Hankinson SE, Hunter DJ. No association between BRCA2 N372H and breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2005;14:1353-1354.

37. The Breast Cancer Linkage Consortium. Cancer Risks in BRCA2 Mutation Carriers. J Natl Cancer Inst. 1999; 91:1310-1316

38. Goto M, Miller RW, Ishikawa Y. Sugano H. Excess of rare cancers in Werner syndrome. Cancer Epidemiol Biomarkers Prev. 1996;5:239-246.

39. Zhou W, Liu G, Miller DP, et al. Polymorphisms in the DNA repair genes XRCC1 and ERCC2, smoking, and lung cancer risk. Cancer Epidemiol Biomarkers Prev. 2003;12:359-365.

40. Nelson HH, Kelsey KT, Mott LA, Karagas MR. The XRCC1 Arg399Gln polymorphism, sunburn, and non-melanoma skin cancer:evidence of gene-environment interaction. Cancer Res. 2002;62:152-155.

41. Smedby KE, Lindgren CM, Hjalgrim H, et al. Variation in DNA repair genes ERCC2, XRCC1, and XRCC3 and risk of follicular lymphoma. Cancer Epidemiol Biomarkers Prev. 2006;15:258-265.

Table 1: Description of NHL Study Participants

Characteristic	Cases		Controls	
	N	%	N	%
	(n=1172)		(n=982)	
Age				
<55	467	39.9	337	34.4
55-64	319	27.2	240	24.4
≥ 65	386	32.9	405	41.2
Race				
White	1006	85.9	787	80.1
African-American	82	7.0	130	13.4
Other	84	7.1	65	4.5
Sex				
Male	643	54.9	516	52.7
Participation:	**76%**		**52%**	
Study site				
Detroit	241	20.5	173	17.6
Iowa	338	28.9	281	28.5
Los Angeles	295	25.2	251	25.6
Seattle	298	25.4	277	28.3
NHL Tumor pathology:				
All B-cell:	955	81.5		
Follicular	280	23.9		
Diffuse Large B-cell	371	31.7		
Small lymphocytic lymphoma	148	12.6		
Marginal Zone*	95	8.1		
Mantle Cell	43	3.7		
Burkitt lymphoma	18	1.5		
All T-cell: ‡	73	6.2		
Not specified:	144	12.3		

* MALT (n=62), Marginal Zone (n=33); ‡ Includes Mycoses Fungoides

Table 2: DNA repair pathways and genetic variants examined in relation to risk of Non-Hodgkin lymphoma.

Gene	Name	Substitution	Location	RS#
NON-HOMOLOGOUS END JOINING/V(D)J				
LIG4	DNA Ligase IV	T91	13q33	rs1805388
RAG1*	Recombination-Activating Gene 1	K820R	11p13	rs2227973
XRCC4	X-Ray Repair Cross-Complementing 4	N298S	5q13	rs1805377
		S307		rs1056503
		A247S		rs3734091
HOMOLOGOUS RECOMBINATIONAL REPAIR				
BRCA2	Breast Cancer 2, early onset	N372H	13q12.3	rs144848
		K1132		rs1801406
		S2414		rs1799955
		Ex27-336A>C		rs15869
		N289H		rs766173
		N991D		rs1799944
NBS1	Nijmegan Breakage Syndrome	E185Q	8q21	rs1805794
TP53	Tumor protein p53	P72R	17p13.1	rs1042522
WRN	Werner Syndrome	V114I	8p12	rs4987236
		M387I		rs1800391
		L1074F		rs2725362
		C1367R		rs1346044
XRCC2	X-Ray Repair Cross-Complementing 2	R188H	7q36.1	rs3218536
XRCC3	X-Ray Repair Cross-Complementing 3	T241M	14q32.3	rs861539
NUCLEOTIDE EXCISION REPAIR				
ERCC1	Excision repair cross-complementing 1	IVS5+33A>C	19q13.2	rs3212961
ERCC2	Excision repair cross-complementing 2	D312N	19q13.3	rs1799793
		K751Q		rs13181
ERCC4	Excision repair cross-complementing 4	P379S	16p13.3	rs1799802
ERCC5	Excision repair cross-complementing 5	D1104H	13q22	rs17655
RAD23B		A249V		rs1805329
XPC	Xeroderma Pigmentosum C	K939Q	3p25	rs2228001
BASE-EXCISION REPAIR				
APEX1	Apurinic/apyrimidinic Endonuclease	D148E	14q11.2	rs3136820
PARP	Poly (ADP-ribose) polymerase	V762A	1q41	rs1136410
XRCC1	X-Ray Repair Cross-Complementing 1	Q399R	19q13.2	rs25487
		R280H		rs25489
		R194W		rs1799782
DIRECT DAMAGE REVERSAL				
MGMT	O-6-methylguanine DNA - methyltransferase	I143V	10q26	rs2308321
		K178R		rs2308327
		L84F		rs12917

* RAG1 is not considered a DNA repair gene, but rather participates in the V(D)J recombination process with DNA repair genes active in NHEJ.

Table 3: Risk of NHL in relation to variants in DNA repair genes[*†‡]

Variant	Cases	Controls	OR 95% CI
Non-Homologous End Joining(NHEJ)/V(D)J Genes			
LIG4 (T9I)			
CC	792	628	1.0
CT	300	275	0.9 (0.7-1.1)
TT	18	28	0.5 (0.3-0.9)
CT or TT			0.8 (0.7-1.0)
p-trend:			.03
RAG1 (K820R)			
AA	814	722	1.0
AG	256	185	1.3 (1.0-1.6)
GG	35	15	2.7 (1.4-5.0)
AG or GG			1.4 (1.1-1.7)
p-trend:			*.0008*
Homologous recombination repair (HR):			
BRCA2 (N372H)			
AA	577	505	1.0
AC	441	361	1.1 (0.9-1.3)
CC	98	60	1.5 (1.0-2.1)
CC or AC			1.1 (0.9-1.3)
p-trend:			*.07*
WRN (V114I)			
GG	652	533	1.0
GA	86	91	0.8 (0.6-1.1)
AA	5	9	0.4 (0.1-1.3)
GA or AA			0.7 (0.6-1.0)
p-trend:			*.04*
Base Excision Repair (BER)			
XRCC1 (R194W)			
CC	916	804	1.0
CT	186	116	1.4 (1.1-1.8)
TT	9	7	1.3 (0.5-3.4)
CT or TT			1.4 (1.1-1.8)
p-trend:			*.0069*

[*] Abbreviations: OR – odds ratio, CI – confidence interval.

[†] All analyses adjusted for reference age (<55, 55-64, ≥ 65), sex, race (White, African-American, other), and study site.

[‡] Participants with missing data omitted.

Table 4: Risk of NHL in relation to variants in DNA repair genes, by NHL case pathology[*†‡§].

	All B-Cell		Follicular		DLBCL		All T-Cell		Controls
	Cases	**OR 95% CI**	**Cases**	**OR 95% CI**	**Cases**	**OR 95% CI**	**Cases**	**OR 95% CI**	
Non-Homologous End Joining/ V(D)J Genes:									
LIG4 (T9I)									
CC	720	1.0	197	1.0	249	1.0	44	1.0	628
CT	267	0.8 (0.7-1.0)	68	0.8 (0.6-1.1)	89	0.8 (0.6-1.1)	24	1.2 (0.7-2.0)	275
TT	17	0.5 (0.3-1.0)	2	NC¶	6	0.6 (0.2-1.4)	1	NC¶	28
CT or TT		0.8 (0.7-1.0)		0.7 (0.5-1.0)		0.8 (0.6-1.0)		1.2 (0.7-2.1)	
p-trend:		*.02*				*.07*			
RAG1 (K820R)									
AA	737	1.0	190	1.0	263	1.0	52	1.0	722
AG	230	1.3 (1.0-1.6)	59	1.3 (0.9-1.8)	75	1.1 (0.8-1.6)	16	1.2 (0.7-2.2)	185
GG	32	2.8 (1.5-5.3)	12	5.1 (2.3-11.7)	5	1.3 (0.2-9.8)	1	NC¶	15
AG or GG		1.4 (1.1-1.7)		1.5 (1.1-2.1)		1.2 (0.7-2.2)		1.2 (0.7-2.2)	
p-trend:		*.0008*		*.0009*		*.33*			
Homologous recombinational repair (HR) :									
BRCA2 (N372H)									
AA	527	1.0	131	1.0	182	1.0	27	1.0	505
AC	391	1.0 (0.8-1.2)	110	1.1 (0.8-1.5)	133	1.0 (0.8-1.3)	35	1.8 (1.1-3.1)	361
CC	89	1.4 (1.0-2.1)	25	1.6 (1.0-2.6)	32	1.5 (1.0-2.4)	9	3.0 (1.3-6.8)	60
CC or AC		1.1 (0.9-1.3)		1.2 (0.9-1.6)		1.1 (0.8-1.4)		2.0 (1.2-3.3)	
p-trend:		*.12*		*.10*		*.21*		*.003*	

Table 4: Risk of NHL in relation to variants in DNA repair genes, by NHL case pathology[*†‡§] (cont).

	All B-Cell		Follicular		DLBCL		All T-Cell		Controls
	Cases	OR 95% CI	Cases	OR 95% CI	Cases	OR 95% CI	Cases	OR 95% CI	
Homologous recombinational repair (HR) :									
WRN (V114I)									
GG	591	1.0	164	1.0	184	1.0	44	1.0	553
GA	78	0.8 (0.6-1.1)	22	0.8 (0.5-1.4)	23	0.7 (0.5-1.2)	6	0.8 (0.3-2.0)	91
AA	5	0.5 (0.2-1.5)	1	NC¶	2	NC¶	0	NC¶	9
GA or AA		0.8 (0.6-1.0)		0.7 (0.5-1.2)		0.7 (0.3-1.8)			
p-trend:		.06							
Base Excision Repair (BER):									
XRCC1_03 (R194W)									
CC	828	1.0	218	1.0	289	1.0	57	1.0	804
CT	167	1.4 (1.1-1.9)	41	1.3 (0.9-2.0)	52	1.3 (0.9-1.8)	13	1.6 (0.8-3.0)	116
TT	9	1.4 (0.5-3.9)	4	NC¶	4	NC¶	0	NC¶	7
CT or TT				1.4 (1.0-2.1)		1.3 (0.9-1.8)		1.5 (0.8-2.8)	
p-trend:		*.007*							

[*] Abbreviations: OR – odds ratio, CI – confidence interval, DLBCL – Diffuse large B-cell lymphoma.

[†] All analyses adjusted for reference age (<55, 55-64, ≥ 65), sex, race (White, African-American, Other), and study site.

[‡] Participants with missing data omitted.

[§] ICD-O codes for tumors included in each subgroup: B-cell lymphoma: 9590,9591,9595,9670-73,9675,9676, 9678-82, 9684, 9686-88, 9690, 9691, 9695-9698, 9710,9711, 9713,9714; Follicular lymphoma: 9676, 9690, 9691,9695-9698; DLBCL: 9678 -9682, 9684, 9688; T-cell lymphoma: 9700, 9702,9705, 9706, 9708, 9709, 9713, 9714.

[¶] NC- Not calculated for cell sizes of 5 or less.